

ASYMMETRY IN VOWEL PERCEPTION IN L1: EVIDENCE FROM ARTICULATORY SYNTHESIS OF AN [i]-[e] CONTINUUM

Charalampos KARYPIDIS¹ - Gilles GUGLIELMI² - Angelica V. COSTAGLIOLA¹⁻³

¹LPP - UMR 7018, CNRS / Université de la Sorbonne Nouvelle- Paris 3, PARIS

²ARP / UFRL Université Paris 7 - Denis Diderot, PARIS

³CRIL / Università degli Studi Lecce - Dip. Filologia, Linguistica e Letteratura, LECCE
ch_karypidis@yahoo.com, gillesgug@yahoo.fr, angelica.costagliola@univ-paris3.fr

1. INTRODUCTION

The phenomenon of asymmetry in vowel perception (also known as "order effect") is first explicitly described by Cowan & Morse (1986), who claim that, in a pair of stimuli, the perceived quality of the first token gradually changes towards a neutral position on the vowel space. Their assumption is based on previous results provided by Shigeno & Fujisaki (1980) who claim that the effect of the second vowel on the labelling of the first (retroactive contrast) is greater than the reverse (proactive contrast), because the first vowel has to be held longer in memory. Nonetheless, results from Experiment 1 of Cowan & Morse (1986) suggest that vowels may each have their own neutral point. Repp & Crowder (1990), citing results of a previous paper (Repp *et alii*, 1979), consider retroactive contrast to provide an implausible explanation for the order effect.

In a very intriguing paper, Medin & Barsalou (1987: 474-475) discuss the notion of "reference points", for both sensory perception and generic knowledge: "Reference points can be either salient values on dimensions that structure categories or they can be prototypes that contain characteristic and ideal attributes of the category.". According to Kuhl (1991), these reference points are none others than the prototypic exemplars of phonemic categories, which she calls "perceptual magnets". By conducting a series of experiments, she provides convincing evidence that vowel categories are structured according to stimulus typicality and that prototypic exemplars play an important role in perception, given that their nature is represented in long-term memory.

On the other hand, a series of articles contest the validity of the asymmetry phenomenon. Most notably, Lotto *et alii* (1998: 3648) argue that the order effect is the mere result of a methodological inconsistency: "[...] category membership is determined by identification of sounds in isolation, whereas, discrimination tasks include pairs of stimuli".. At the same time, they support that the choice of stimuli made by Kuhl was not appropriate since her non-prototypic [i] actually belonged to another phonemic category. This last counterargument is also supported by Sussman & Lauckner-Morano (1995).

Based on a number of experiments carried out by various researchers (Kuhl, 1991; Swoboda *et alii*, 1978, among others), Polka & Bohn (2003) generalise Kuhl's assumption, claiming that virtually any vowel can serve as a reference point, provided it is more peripheral than the following vowel: "Asymmetries in vowel perception occur such that discrimination of a vowel change presented in one direction is easier compared to the same change presented in the reverse direction [...] the more peripheral vowel within a contrast serves as a reference or

perceptual anchor.”. Nonetheless, the term “peripheral” has been subject to many interpretations. Traditionally, it is linked to either an extreme position on one of the corners of the vowel space or a position on the periphery (edges) of the vowel space. In this paper, we consider as “peripheral” any token located on one of the corners of the vowel space (be it on the F1-F2, F2-F3 or F1-F2' dimension).

In previous studies, vowel continua were synthesized by fragmenting, in equidistant points, the F1/F2 Euclidean distance in Hertz between two prototypes (best representatives of two phonemic categories). This method yielded unrealistic intermediary sounds, inasmuch as most of them were assigned formant-value combinations that cannot be produced by a human vocal tract (Boë *et alii*, 1989). Furthermore, all members of the [i]-[e] continuum (upon which this paper focuses) were assigned fixed F3 and F4 values (3010 and 3300 Hz respectively), generating a false spectral peak grouping F3 and F4 – a typical attribute of [i] in French (see Table 1) and American English (Hillenbrand, 1995). This, in turn, induced listeners to identify more [i]'s than they should have and thus shifted the identification boundary towards [e] (for similar results, see Iverson & Kuhl, 2000). This proximity of F3 and F4 was emphasized by the selected cascade formant synthesizer (Klatt, 1980), which reinforced the amplitude of the two formants, overshadowing thus perceptively the lower formants, upon which the aforementioned papers focused; a parallel synthesizer, on the contrary, would treat the two formants as discrete and would not take their relatively minimal distance into account. Moreover, evidence from a recent study on vowel prototypes (Karypidis *et alii*, in preparation-b) suggests that [i] has a very narrow perceptual zone, despite its acoustic stability (Stevens, 1989) and notwithstanding the absence of neighbouring vowels (mid-close [e] or [y]) in the system, which would potentially take up some of its perceptual space.

Our initial hypothesis being that the perception of the [i]-[e] contrast is not mainly based on F1 or F2, but on F2', we have examined the extent of the phenomenon of asymmetry (that is, whether discrimination between two vowels is indeed facilitated when the more peripheral vowel is presented second) by constructing a vowel continuum where all acoustic parameters were modified simultaneously. The only method for that to be achieved in a naturalistic way is, as far as we know, by using an articulatory model such as VTCALCs (Maeda, 1990), which “offers the advantage of physiological realism by integrating articulatory constraints” (Boë *et alii*, 1989). By modifying all parameters simultaneously, listeners would be obliged to use the whole spectrum of each stimulus to find acoustic cues instead of solely trying to extract F1 and F2 from the spectrum (which was the case in previous studies, since F3, F4 and F5 were always fixed).

2. EXPERIMENT 1

In this experiment, the listeners identified each stimulus as [i] or [e]. The aim of this study was to locate the [i]-[e] identification boundary and to examine which formants allow the listeners to distinguish the two phonemes. The location of this boundary was later used to interpret the results of Experiment 2.

2.1 Participants

Thirty-four French subjects from 18 to 51 years of age (mean=30.9 years; standard deviation=9 years) participated in the experiment. All reported being native speakers of French

and having no known hearing impairments.

2.2 Stimuli

The stimuli were synthesized with Maeda's articulatory model (1990), which was run on a Toshiba Satellite A10 computer.

Stimulus	Hertz				Bark ¹				
	F1	F2	F3	F4	F1	F2	F3	F4	F2'
[i]	308	2064	2976	3407	3.11	13.22	15.63	16.49	15.92
[e]	365	1961	2644	3362	3.68	12.88	14.87	16.41	13.54
1	247.6	2290.5	3157.1 ²	3838.1	2.4	13.92	16.01	17.22	16.41
2	278.6	2290.5	3157.1	3838.1	2.7	13.92	16.01	17.22	16.41
3	309.5	2259.5	3002.4	3838.1	3	13.83	15.69	17.22	16.20
4	309.5	2228.6	2909.5	3838.1	3	13.73	15.49	17.22	16.07
5	340.5	2228.6	2847.6	3838.1	3.3	13.73	15.35	17.22	14.27
6	340.5	2197.6	2785.7	3807.1	3.3	13.64	15.21	17.17	14.16
7	371.4	2166.7	2723.8	3807.1	3.6	13.55	15.06	17.17	14.05
8	371.4	2135.7	2661.9	3807.1	3.6	13.45	14.91	17.17	13.94
9	402.4	2104.8	2631	3807.1	3.9	13.35	14.83	17.17	13.84
10	402.4	2042.9	2600	3776.2	3.9	13.15	14.76	17.12	13.69

Table 1: Acoustic parameters in Hertz and Bark¹ for the French [i] and [e] of male speakers (Calliope, 1989: 84) and for the ten synthesized stimuli.

We started by synthesizing a prototypic³ [i] (stimulus no. 1, the most peripheral) using the appropriate .LAM (Linear Articulatory Model) file provided by the application. We then modified the jaw height and tongue position parameters (the only articulatory differences between the two vowels, according to the model) in 9 gradual steps toward a prototypic³ [e] (stimulus no. 10, the least peripheral). We thus obtained a 10-vowel continuum. For all stimuli, fundamental frequency rose from 121 to 130 Hz over the first third and then declined to 100 Hz over the final two-thirds. Duration was fixed at 350 ms for all tokens. The stimuli were synthesized at an 11025-Hz sampling rate and at a 16-bit depth and were saved in mono .wav files and in PCM (Pulse Code Modulation) format. Formantic values for the 10 stimuli can be

¹ A Hertz-to-Bark converter is available at: <http://www.ling.su.se/staff/hartmut/umrechnung.htm>

² VTCALCs is unable to calculate the actual F3 because of its very low amplitude or bandwidth. In this case, we have considered F3 of stimulus 1 to be equal to F3 of stimulus 2.

³ We have considered stimuli no. 1 and 10 as prototypic based on 3 criteria: a) their acoustic similarity to the prototypes proposed in the literature (cf. Table 1), b) their near-perfect identification accuracy and c) the magnitude of perceptual warping effects accounted for around them (the closer a token to a prototype, the harder the discrimination).

found in Table 1.

2.3 Procedure

The tokens were presented over Creative headphones in a small, quiet room and the experiment was run on a Toshiba 300CDS laptop by the second author of this paper. An approximate sound-pressure level was chosen intuitively in order for the tokens to give a realistic impression (that is, as naturalistic as synthesized vowels could sound) and was identical for all participants. The software which served as interface for the experiment was Praat for Windows (Boersma & Weenink, 2001). All stimuli were presented seven times each, in random order and without doublets (consecutive repetitions of the same stimulus). The two possible answers were *i* and *é* (undeniable orthographical equivalents of [i] and [e] in French), presented on the screen in yellow squares, on which listeners were asked to click with a mouse to give an answer. After each reply, the following stimulus was presented with a 0.5-second delay. There was a break every 20 stimuli and listeners were requested to click anywhere on the screen to continue. Most listeners clicked through this screen message almost instantly, an indication that the experiment was not cumbersome.

The experiment was preceded by a short training session where all 10 initial vowels were presented twice, in random order and without doublets.

2.4 Results - Discussion

First of all, the need to examine whether the identification rate for each stimulus was above chance was apparent and a binomial test (Uitenbroek, 1997)⁴ was conducted. Therefore, for $N=238$ (7 repetitions of each stimulus x 34 subjects), $\alpha=0.01$ (level of confidence interval) and $\pi=0.5$ (expected proportion of either response: [i] or [e]), the identification rate for each stimulus is above chance (>59%).

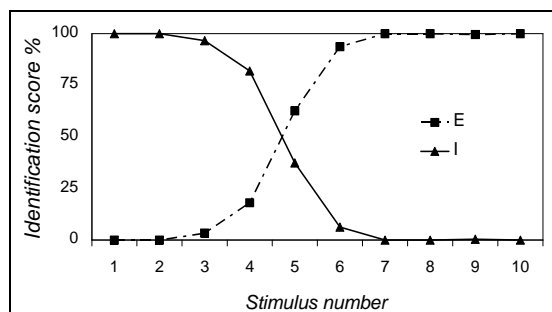


Figure 1: Identification scores for the ten stimuli.

A graphical representation of the identification results (Figure 1) indicates that the transition from [i] to [e] is highly quantal. Stimulus 4 has the lowest rate of all other members of the [i] category even though its formants are quite close to those of a production prototype [i] proposed in the literature (Table 1) whilst stimuli 1 to 3 were much fronter than expected. The

⁴ Available on: <http://home.clara.net/sisa/binomial.htm>

[e] category was not an exception in our study: all stimuli (6-10) identified as [e] at a very high rate (>90%) were far more front than expected, with their F2 being much higher than that of a prototypic uttered French [e].

We have conducted an ANOVA test (using the statistics software Systat 11.0) with a single factor (stimulus) to verify whether the listeners respond the same way to all stimuli and a significant effect of stimulus on response was found [$F(9) = 406.2711$, $p < 0.01$].

On the other hand, Figure 1 depicts a sudden change in labelling, which is rather unusual for vowels, not to mention long vowels. Moreover, formant extraction (Table 1) cannot sufficiently explain this abrupt change from [i] to [e], since formants follow a somewhat continuous evolution (non-quantal articulatory/acoustic relationship). Therefore, we converted these values into Bark using Traummüller's formula (1990)¹:

$$z = [26.81 / (1 + 1960/f)] - 0.53$$

where f stands for "frequency" and z for "critical band rate z " (in bark) and have calculated $F2'$ (Mantakas, 1986)⁵ for each stimulus (Table 1, last column). This time, the non-linearity in the evolution of $F2'$ (quantal relationship between acoustic output and perception) explains perfectly the identification results since the tokens identified as [i] are distinguished from the [e] members by the presence of a [+extreme high $F2'$] acoustic cue. Consequently, we postulate that $F2'$ plays a certain role in the [i]-[e] contrast in French, at least in isolated context.

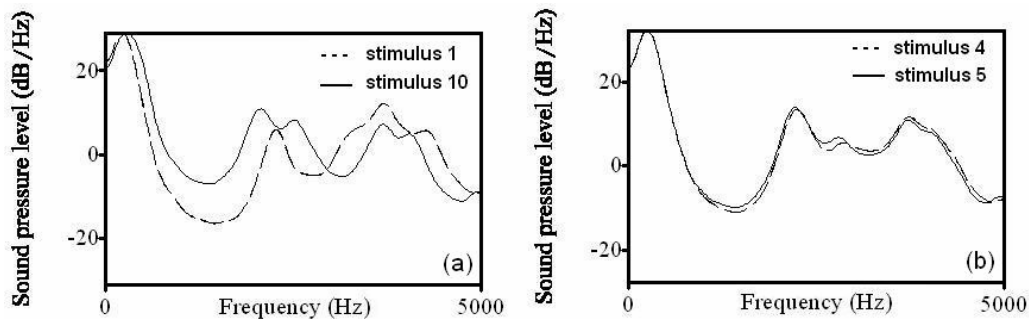


Figure 2. Superimposed amplitude spectra of a) the prototypic [i] and [e] (stimuli 1 and 10) and b) the two tokens located near the identification boundary (stimuli 4 and 5). Dashed lines represent vowels identified as [i] and plain lines depict [e] members.

To better understand the usefulness of $F2'$ (which represents the overall perceptual weight of higher formants – F2, F3 and F4 – and whose calculation is dependent on their relative

⁵ The formula is also available in Schwartz (1997). A Praat script is available on the first author's Web page: http://www.geocities.com/ch_karypidis/

distance), we prepared amplitude spectra (Figures 2a and 2b) of four tokens: stimuli 1, 4, 5 and 10 with the acoustics software Praat (Boersma & Weenink, 2001). For the extraction of the spectral slices, we have selected a 0.0389-ms window (the 5 periods between the 12th and 17th pulse, that is from the 12th to the 16th period) used. The starting point of the selection was at 0.09577 ms from the start of the file and the ending point at 0.13457 ms. Mean pitch was of 128.67 Hz and the middle point of the window roughly corresponded to the highest pitch point of the whole token. The window shape was Gaussian and a 6-dB/oct pre-emphasis was chosen. After the spectral extraction, we have used a cepstral smoothing (bandwidth=500Hz) to render the spectra more legible. The 4 tokens aforementioned were chosen either because they are prototypic (stimuli 1 and 10) or because they are located on the boundaries of the two categories (stimuli 4 and 5), perceived as such in Experiment 1.

Figure 2a depicts a large energy concentration at around 3800 Hz for stimulus 1, whilst a significant amount of energy for stimulus 10 was located around 2200 Hz. This difference of energy dislocation is, in fact, responsible for the discrepancy in F2' between the two prototypes (Table 1). Having this in mind, we have examined Figure 2b in order to understand the sudden change of answering pattern. A closer look evinces that the spectral difference between the two tokens perceived as belonging to distinct categories is hardly noticeable [Flanagan (1957) reports that the minimum precision to quantize formant amplitude is about ± 3 dB for F2 and ± 5 dB for F3]. Thus, the location of the boundary between [i] and [e] cannot be explained on spectral/acoustic grounds.

Evidently, a lot of questions arise from our results and methodology and further experimentation is needed lest the role of F2' be elucidated. An ongoing research on the [i]-[y] contrast tries to re-examine the dislocation of energy packs and the quantal nature of the relationship between acoustic output and perception. In addition, an investigation of the [u]-[o] and [y]-[ø] contrasts would allow us to verify whether F2' (either high or low) is linked to the [+high] feature in French.

3. EXPERIMENT 2

This experiment consisted of an AX discrimination task, where the stimuli of Experiment 1 were presented in pairs and in both orders, while differing in 1 or 2 steps along the continuum. The aim of this study was to examine whether the order effect can be accounted for when using articulatory synthesis, thus when modifying all formants at the same time. In addition, we have tried to verify whether the order effect occurs only when extreme vowels are concerned or whether it is linked to typicality.

3.1 Participants

Listeners are as described in Section 2. One participant was excluded for having judged all pairs as different (mean=30.6 years; standard deviation=8.9 years).

3.2 Stimuli

Stimuli are as described in Section 2.

3.3 Procedure

External conditions were as described in Experiment 1. Listeners were presented with 5

repetitions of 18 one-step pairs (9 stimulus combinations, 2 orders) and 16 two-step pairs (8 stimulus combinations, 2 orders), in random order and without doublets. In the Forward order, the first stimulus presented was more extreme than the second whilst in the Reverse order, the most extreme token was presented second. Table 2 demonstrates the 4 different stimulus combinations:

Group of pairs	Order	Step difference	Label	Number of pairs
1-2, 2-3, ... , 9-10	Forward	One	For1	9
2-1, 3-2, ... , 10-9	Reverse	One	Rev1	9
1-3, 2-4, ... , 8-10	Forward	Two	For2	8
3-1, 4-2, ... , 10-8	Reverse	Two	Rev2	8

Table 2: The four groups of pairs (“Label”=the way the group of pairs is referred to throughout this paper).

The Inter-Stimulus Interval (ISI) was fixed at 250 ms. According to Cowan & Morse (1986), information stocked in the auditory memory is most efficiently preserved when the ISI is fixed at this value. Subjects were asked whether the vowels presented in pairs were the same or different. After each answer, the first vowel of the following pair was presented with a 0.5-second delay. The experiment was interrupted every 15 pairs of stimuli, proposing a short break and asking listeners to click anywhere on the screen to continue. Listeners admitted that this experiment was more cumbersome than Experiment 1.

The experiment was preceded by a short training session where all 34 pairs were presented once and in random order.

3.4 Results and discussion

The order effect predicts that mean discrimination scores (M) for pairs where the most peripheral element is presented first (For1 and For2) would be lower than for pairs in the opposite order (Rev1 and Rev2 respectively). In other words, our hypotheses (H) are:

$$H_0: M_{For1} < M_{Rev1}$$

$$H_1: M_{For1} < M_{Rev1}$$

$$H_2: M_{For2} < M_{Rev2}$$

Paired samples t-tests reveal that the order effect is significant for both H_1 [$t(296) = -5.7368$, $p < 0.01$] and H_2 [$t(263) = -4.9795$, $p < 0.01$], thus H_1 and H_2 are valid.

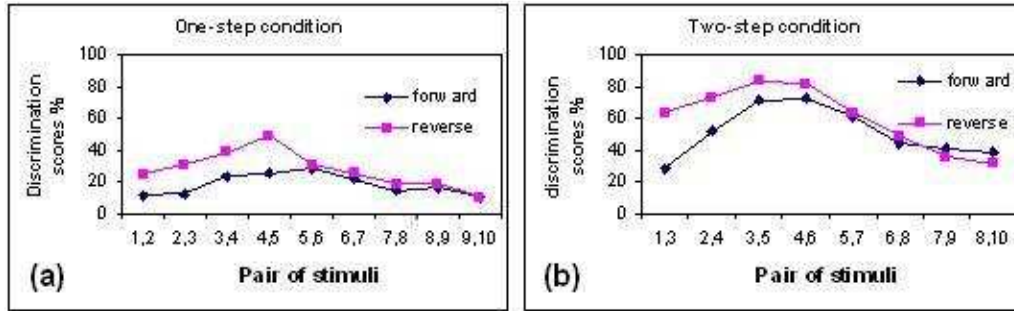


Figure 3. Discrimination scores for pairs in a) one-step and b) two-step condition.

Nonetheless, the line graphs in Figures 3a and 3b suggest that, in both conditions (one- and two-step), the only pairs accounting for the asymmetry phenomenon are those in which the most peripheral stimulus is characterised by an extremely high F2' (that is, the pairs containing stimuli 1, 2, 3 or 4, all identified as [i]). Additional paired samples t-tests verified this assumption (Table 3).

Group of pairs of vowels		Degree of Freedom	t value	p value
A	1-2, ..., 4-5	131	-6.8017	<0.01
B	5-6, ..., 9-10	164	-1.3286	0.09
C	1-3, ..., 4-6	131	-7.5669	<0.01
D	5-7, ..., 8-10	131	0.3681	0.64

Table 3: Results of paired samples t-tests (Forward vs. Reverse Order) for pairs with at least one [i] element (groups A, C) and pairs with no [i] elements (groups B, D). For each group, scores for Forward and Reverse order have been compared.

The results of Table 3 (that the asymmetry phenomenon is triggered only by [i] members), could be explained by the fact that, as mentioned in the Introduction, [i] is considered to be an acoustically more stable vowel than [e] (Stevens, 1989; Badin *et alii*, 1990). Schwartz *et alii* (2005), citing a previous article (Schwartz & Escudier, 1989), report that stimuli with formant convergence (either F2-F3 or F3-F4) are more stable in short-term memory, producing a level of false alarms lower than do non-focal tokens. Evidently, an experiment focusing on the magnitude of the ISI would clarify the role the auditory and phonetic memory play in discrimination and in the mechanism of asymmetry. According to Repp (1990), at long ISIs listeners rely more on the phonemic categorization of the first vowel to discriminate, while, at short ISIs, they mostly use their auditory memory.

We also attempted to verify whether the magnitude of acoustic difference (step condition) has an effect on discriminability. For that purpose, we made the following comparisons: a) For1 vs. For2 and b) Rev1 vs. Rev2. Two-sample t-tests indicate that listeners discriminate better in the two-step condition: a) For1 vs. For2: [t(559)= -14.27, p<0.01] and b) Rev1 vs. Rev2: [t(559)= -12.44, p<0.01], that is, when acoustic difference is greater. The discrimination curves in Figures 3a and 3b demonstrate a downward slope, with the peaks

coinciding with the pairs containing elements from two different phonemic categories (pairs 4-5, 3-5 and 4-6). According to Lotto (1998), the location of this peak should not surprise us since cross-boundary contrasts generally yield higher discrimination rates. In addition, the shape and location of these curves conform with Kuhl's (1991) results: "When a stimulus perceived as having high category goodness was used as the referent vowel in the discrimination task, overall percent-correct scores were significantly lower, indicating difficulty in perceiving differences between the prototype and other members of the category.". Nonetheless, unless the prototype of the [e] category is not located in the range of our synthesized continuum (however, identification scores for stimuli 7 to 10, are close to 100%), Kuhl's hypothesis cannot explain why the order effect occurs only with the [i] members. In addition, it is not quite clear how the information concerning the level of typicality ("good"/"bad" exemplar) is stored in our short-term memory.

In Figure 4, we have calculated the acoustic difference (for all 4 formants) between stimuli pairs. We can see again that discrimination peaks of Figures 3a et 3b do not coincide with peaks in Figures 4a and 4b. and that it is, once again, the F2' hypothesis that provides a better account of peaks.

Moreover, discrimination scores for pair 4-5 (the peak in the one step-condition, Reverse Order) does not meet the level of chance (score=49%), suggesting that the acoustic difference (separate formants and relative amplitude) between the two tokens is infinitesimal and that discrimination for this pair is not mainly based on phonetic labelling.

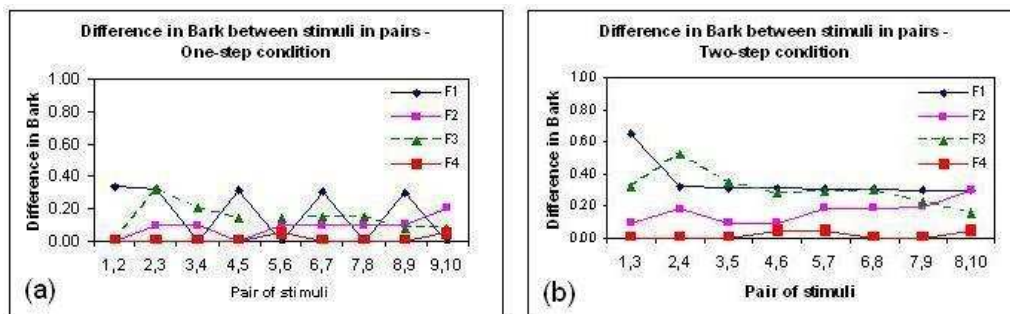


Figure 4. Differences in Bark among the paired stimuli.

4. CONCLUSION

Polka & Bohn (2003) have postulated that extreme vowels have the ability to remain in short-term memory longer than non-extreme tokens. We have used articulatory synthesis (Maeda, 1990) to prepare an [i]-[e] continuum and have studied the importance of higher formants for the aforementioned contrast in French.

According to literature (Calliope, 1989), a prototypic uttered French [i] is focal, displaying a proximity of F3 and F4 (and thus a very high F2'). Indeed, identification results followed the abrupt evolution of F2' along the continuum: when F3 was too far from F4 to form a single energy pack (which would amplify their respective intensity), responses equally changed to [e],

indicating that [i] is expected to have an extremely high F2. On the other hand, relative formant amplitude and separate formant evolution did not converge with identification curves.

Experiment 2 has provided evidence against Polka's generalizing hypothesis, given that the phenomenon of asymmetry occurred only for our [i] stimuli. If reference points were located at phoneme boundaries, both phonetic categories would demonstrate an order effect. A possible explanation was given by taking the focal nature (proximity of F3 and F4) of [i] into account. Based on results provided by Schwartz & Escudier (1989), we have assumed that focal vowels are more stable in the short-term memory, which makes discrimination more difficult. Further experiments with other focal vowels ([u], [a], [y] and [o]) would elucidate their role in our perception. Once again, acoustic difference between paired stimuli did not converge with discrimination curves.

Results from an ongoing research on Southern Salentinian Italian (SSI) and Spanish (Karypidis *et alii*, in preparation-a) indicate that the order effect occurs only in systems where [i] and [e] are characterised by discretely distinct F2' values. More specifically, in SSI, only [i] has a very high F2' (Grimaldi, 2003) whilst both Spanish [i] and [e] are extreme on the F2' dimension (Quilis & Esgueva, 1983). This latest evidence supports our hypothesis that, in certain languages, the [i]-[e] contrast could be mostly based not on the difference in F1 or F2 but on the presence or absence of a [+extreme high F2] acoustic cue.

On the other hand, we could postulate with a certain confidence that [i] members are characterized by an onward slope (starting from F2) whereas [e] members demonstrate a downward slope. We assume thus that this acoustic parameter (energy concentration at different location in the higher frequencies) is behind the discriminability of stimuli 4 and 5.

5. ACKNOWLEDGEMENTS

The authors are grateful to Nick Clements, Annie Rialland and Mirko Grimaldi for commenting on different versions of this paper as well as 2 anonymous reviewers. Thanks are extended to all 34 participants who lent us their time and energy.

6. BIBLIOGRAPHY

- Badin, P., Perrier, P., Boë, L.-J. & Abry, C. (1990), Vocalic nomograms: Acoustic and articulatory considerations upon formant convergences, *Journal of the Acoustical Society of America*, 87, 1290-1300.
- Boë, L. J., Perrier, P., Guérin, B. & Schwartz, J. L. (1989), Maximal vowel space, in *Proceedings of the First European Conference on Speech Communication and Technology Paris*, 281-284.
- Boersma, P. & Weenink, D. (2001), PRAAT, a system for doing phonetics by computer, *Glott International*, 5(9/10): 341-345.
- Calliope (1989), *La parole et son traitement automatique* (J. P. Tubach), Paris, Masson.
- Cowan, N. & Morse, P. A. (1986), The use of auditory and phonetic memory in vowel discrimination, *Journal of the Acoustical Society of America*, 79, 500-507.

Flanagan, J. L. (1957), Estimates of the Maximum Precision Necessary in Quantizing Certain "Dimensions" of Vowel Sounds, *The Journal of the Acoustical Society of America*, 29(4), 533-534.

Grimaldi, M. (2003), *Nuove ricerche sul vocalismo tonico del Salento meridionale. Analisi acustica e trattamento fonologico dei dati*, Alessandria, Edizioni dell'Orso.

Hillenbrand, J., Getty, M., Clark, M. & Wheeler, K. (1995), Acoustic characteristics of American English vowels, *Journal of the Acoustical Society of America*, 97, 3099-3111.

Iverson, P. & Kuhl, P. K. (2000), Perceptual magnet and phoneme boundary effects in speech perception: do they arise from a common mechanism?, *Perception & Psychophysics*, 62, 874-886.

Karypidis, Ch., Colazo-Simon, A. & Costagliola, A. V. (in preparation-a), Asymmetry in vowel perception in L2: evidence from articulatory synthesis of an [i]-[e] continuum.

Karypidis, Ch., Costagliola, A. V. & Colazo-Simon, A. (in preparation-b), Vowel prototypes assimilation: a cross-linguistic perceptual study of five-vowel systems.

Klatt, D. H. (1980), Software for a cascade/parallel formant synthesizer, *Journal of the Acoustical Society of America*, 67, 737-793.

Kuhl, P. K. (1991), Human adults and human infants show a "perceptual magnet effect" for the prototypes of speech categories, monkeys do not, *Perception & Psychophysics*, 50, 93-107.

Lotto, A. J., Kluender, K. R. and Holt, L. L. (1998), Depolarizing the perceptual magnet effect, *Journal of the Acoustical Society of America*, 103, 3648-3655.

Maeda, S. (1990), Compensatory articulation during speech; evidence from the analysis and synthesis of vocal-tract shapes using an articulatory model, in *Speech Production and Speech Modelling*, NATO ASI Series (A. Marchal & W. J. Hardcastle), Kluwer Academic Publishers, 131-149.

Mantakas, M., Schwartz, J. L. & Escudier, P. (1986), Modèle de prédiction du 'deuxième formant effectif' F₂ - application à l'étude de la labialité des voyelles avant du français., in *Proceedings of 15th Journées d'étude sur la parole*, 157-161.

Medin, D. L. & Barsalou, L. W. (1987), Categorical processes and categorical perception, in *Categorical Perception* (S. Harnad, editor), Cambridge: Cambridge University Press, 455-490.

Polka, L. & Bohn, O.-S. (2003), Asymmetries in vowel perception, *Speech Communication*, 41, 221-231.

Quilis, A. & Esqueva, M. (1983), Realización de los fonemas vocálicos españoles en posición fonética normal, in ESGUEVA, M.- CANTARERO, M. (Eds) *Estudios de Fonética I*, Madrid: CSIC (Collectanea Phonetica VII), pp. 137-252.

Repp, B. H. & Crowder, R. G. (1990), Stimulus order effects in vowel discrimination, *Journal of the Acoustical Society of America*, 88(5), 2080-2090.

Repp, B., Healy, A. F. & Crowder, R. G. (1979), Categories and context in the perception of isolated steady-state vowels, *Journal of Experimental Psychology: Human Perception and Performance*, 5, 129-145.

Schwartz, J.-L., Abry, C., Boë, L.-J., Menard, L. & Vallée, N. (2005), Asymmetries in vowel perception, in the context of the Dispersion-Focalisation Theory, *Speech Communication*, 45, 425-434.

Schwartz, J.-L. & Escudier, P. (1989), A strong evidence for the existence of a large-scale integrated spectral representation in vowel perception, *Speech Communication*, 8, 235-259.

Shigeno, S. & Fujisaki, H. (1980), Context Effects in Phonetic and Non-Phonetic Vowel Judgments, *Annual Bulletin Research Institute of Logopedics and Phoniatrics*, Faculty of Medicine, University of Tokyo, 14, 217-224.

Stevens, K. N. (1989), On the Quantal Nature of Speech, *Journal of Phonetics*, 17, 3-45.

Sussman, J. E., and Lauckner-Morano, V. J. (1995), Further tests of the perceptual magnet effect in the perception of [i]: Identification and change/no-change discrimination, *Journal of the Acoustical Society of America*, 97, 539-552.

Swoboda, P. J., Kass, J., Morse, P. A. & Leavitt, L. A. (1978), Memory factors in vowel discrimination of normal and at-risk infants, *Child Development*, 49, 332-339.

Traunmüller, H. (1990), Analytical expressions for the tonotopic sensory scale, *Journal of the Acoustical Society of America*, 88, 97-100.

Uitenbroek, D. G. (1997), SISA-Binomial, <http://home.clara.net/sisa/binomial.htm> (10 February 2004).